Comparing Machine Learning and Neuro-Fuzzy Models for Hydropower Prediction Under Seasonal Operating Conditions

[1]Daniel Rohi,[2]Akhmad Solikin
[1,2]Department of Electrical Engineering, PGRI Adi Buana University, Surabaya, Indonesia
rohi@adibuanasby.ac.id

Abstract - Accurate prediction of hydropower energy production is essential for operational planning and decision-making under varying hydrological conditions. However, comparative evaluations of high-accuracy data-driven models and interpretable intelligent models using real operational data remain limited. This study aims to compare the performance of Machine Learning and Neuro-Fuzzy approaches for predicting energy production at the Sengguruh Hydropower Plant, Indonesia. One year of engineering-validated operational data consisting of discharge, reservoir and tailrace water levels, operation time, turbine efficiency, and derived hydraulic head were used for model development and evaluation. Three Machine Learning models, namely Artificial Neural Network (ANN), Random Forest (RF), and Support Vector Regression (SVR), were compared with an Adaptive Neuro-Fuzzy Inference System (ANFIS). Model performance was evaluated using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2). The results indicate that Machine Learning models achieved superior predictive performance, with Random Forest providing the highest accuracy (RMSE = 8.42 MWh/day, MAE = 6.35 MWh/day, and R^2 = 0.982). In contrast, the ANFIS model produced lower prediction accuracy but offered a more interpretable representation of hydropower operating behavior. Correlation analysis further indicates that operation time and inflow discharge are the most influential operational variables associated with energy production. The findings demonstrate that Machine Learning approaches are more suitable for high-accuracy forecasting and operational optimization, whereas Neuro-Fuzzy models provide advantages in interpretability and operational representation. These results contribute to the development of intelligent hydropower forecasting systems that balance predictive performance and engineering understanding.

Keywords — Hydropower, Machine Learning, Neuro-Fuzzy, Random Forest, seasonal hydrology, operational modeling.

I. Introduction

This paper investigates the application of Machine Learning and Neuro-Fuzzy approaches for hydropower energy prediction using real operational data from the Sengguruh Hydropower Plant.

Hydropower generation is fundamentally governed by hydraulic and energy conversion principles. The power output depends on discharge and head, which are the dominant variables in energy production [21]. The power output of a hydropower plant can be expressed as:

$$P=\rho gQH\eta \quad (1)$$

Where P = generated power (W), ρ = water density (kg/m^3), g = gravitational acceleration (9.81 m/s^2), Q = water discharge (m^3/s), H = net head (m) and η = overall system efficiency. This equation highlights that hydropower output is directly influenced by discharge and head, which are the dominant variables in energy production.

In practice, however, system efficiency varies with operating conditions, and hydrological inputs are highly dynamic. As a result, purely physics-based models are often insufficient to represent real operating behavior, particularly under nonlinear and uncertain conditions.

To overcome the limitations of conventional models, Machine Learning (ML) approaches have been widely adopted in hydropower modeling. These methods learn the relationship between input variables and energy output directly from data without requiring explicit physical equations, enabling effective representation of complex nonlinear system behavior.

Recent studies have demonstrated that ML models achieve high predictive accuracy in hydropower systems [1], [2], [7], [15]. In particular, ensemble learning methods such as Random Forest and gradient boosting are highly effective in handling nonlinear and multivariate data [4], [11], [13], [14], [16]. These models are capable of capturing complex interactions among key variables, including discharge, head, and operation time.

More recent developments have explored advanced architectures such as hybrid CNN–SVR models and deep reinforcement learning for hydropower prediction and optimization [17], [18]. In addition, several studies highlight the growing role of Machine Learning in water–energy systems and smart hydropower management [5], [16], [19], [24]. In general, the ML modeling process can be represented as:

$$y=f(Q,H,t) \quad (2)$$

where y = predicted energy output, Q = discharge, H = net head, t = operation time f(.) = nonlinear function is a nonlinear function learned from data. This data-driven formulation enables flexible modeling but often lacks physical interpretability.

Neuro-Fuzzy models, particularly Adaptive Neuro-Fuzzy Inference Systems (ANFIS), combine fuzzy logic and neural networks to model nonlinear systems through interpretable rule-based structures. The ANFIS model was originally introduced by Jang [8] and has been widely applied in energy prediction problems.

Previous studies have demonstrated the effectiveness of ANFIS in hydropower modeling, particularly in representing nonlinear relationships between discharge, head, and energy production [9], [10]. In addition, earlier work on the Sengguruh Hydropower Plant using Neuro-Fuzzy networks further confirms the capability of this approach in modeling real operational data [26]. A typical fuzzy rule can be expressed as:

"IF Q is High AND H is Medium THEN P is High."

This rule-based structure provides interpretability, allowing engineers to understand the relationship between variables. However, the performance of Neuro-Fuzzy models depends on the selection of membership functions and rule sets, which may limit scalability for complex datasets.

Recent developments in hydropower modeling emphasize the integration of physical knowledge into data-driven approaches, commonly referred to as physics-informed machine learning [6]. This paradigm combines the strengths of physical models and Machine Learning to improve model reliability and consistency. In this framework, the learning process is guided by physical constraints, such as the fundamental hydropower relationship (P \propto QH), which links energy production to discharge and head. Incorporating such relationships enables better generalization and ensures that model predictions remain physically consistent.

In addition, ensemble and hybrid approaches have been applied in water–energy systems to further enhance prediction robustness and reliability [14], [23], [25]. This integrated approach addresses a key limitation of purely data-driven models, namely their lack of interpretability and physical meaning.

Despite significant advances in both Machine Learning and Neuro-Fuzzy modeling, several limitations remain in current hydropower prediction studies. Most existing works focus on a single modeling approach without providing a systematic comparison between high-accuracy data-driven models and interpretable models. In addition, physically meaningful variables, particularly hydraulic head, are often not explicitly incorporated, reducing the physical consistency of the models. Furthermore, many studies rely on synthetic or limited datasets, with insufficient validation using real operational data under varying conditions. The trade-off between predictive accuracy and model interpretability therefore remains insufficiently explored in hydropower energy forecasting.

To address these gaps, this study proposes a comparative framework that integrates physical understanding with data-driven modeling. The main contributions of this study are threefold: (1) the use of real operational data from a hydropower plant, ensuring practical relevance; (2) the explicit incorporation of physically meaningful variables, particularly discharge and head; and (3) a systematic comparison between Neuro-Fuzzy and multiple Machine Learning models to evaluate the trade-off between accuracy and interpretability.

The Sengguruh Hydropower Plant represents a regulated hydropower system operating under substantial seasonal discharge variability while maintaining relatively stable hydraulic characteristics. These conditions provide a suitable case study for evaluating the capability of intelligent prediction models under realistic operational constraints.

The objective of this study is to evaluate the capability of Machine Learning and Neuro-Fuzzy models in predicting hydropower energy production under realistic seasonal operating conditions. The performance of the developed models is assessed using operational data from the Sengguruh Hydropower Plant and evaluated in terms of predictive accuracy, robustness, and interpretability. The results are expected to support operational forecasting, decision-making, and the development of intelligent hydropower management systems.

II. Research Method

A. Research Framework

A systematic research framework is essential to ensure that hydropower energy prediction is developed in a manner that is both physically meaningful and data-driven. Hydropower generation is governed by complex interactions among hydraulic, operational, and environmental variables, requiring modeling approaches capable of representing nonlinear system behavior while maintaining consistency with fundamental engineering principles. Therefore, this study adopts an integrated framework that combines physical hydropower knowledge, Neuro-Fuzzy modeling, and Machine Learning approaches within a unified prediction architecture.

The framework is designed to evaluate the capability of interpretable and data-driven models in predicting hydropower energy production under realistic seasonal operating conditions. The methodology begins with the collection of operational data from the Sengguruh Hydropower Plant, followed by data validation, preprocessing, and feature preparation. Physically meaningful variables, particularly discharge and hydraulic head, are incorporated to preserve the physical characteristics of hydropower operation and improve model reliability.

Subsequently, hydropower energy prediction models are developed using both Neuro-Fuzzy and Machine Learning approaches. The Neuro-Fuzzy model provides an interpretable rule-based representation of system behavior, whereas Machine Learning models learn nonlinear relationships directly from operational data. The developed models are then trained, tested, and evaluated using statistical performance metrics to assess their predictive capability, robustness, and reliability.

The proposed framework enables a systematic comparison between prediction accuracy and model interpretability while maintaining consistency with the physical behavior of hydropower systems. The overall research framework adopted in this study is illustrated in Figure 1.

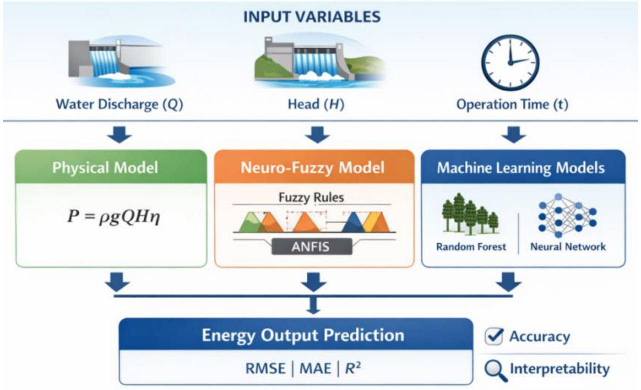

Figure 1. Integrated framework for hydropower energy prediction models

B. Study Area

The Sengguruh Hydropower Plant is located in the upstream area of the Brantas River in Malang Regency, Indonesia, and plays an important role in regional water regulation and electricity generation. The plant has an installed capacity of approximately 29 MW and operates by utilizing water discharge and hydraulic head to convert potential energy into electrical energy, as illustrated in Fig. 2.

The upstream water level ranges between HWL 293.10 m and LWL 291.40 m, while the downstream level is approximately TWL 267.60 m, resulting in an effective head of about 24–26 m. Water flows from the intake through the penstock to drive the turbine and generator, forming a continuous energy conversion process.

These operating conditions indicate that energy production is primarily governed by discharge and hydraulic head, both of which are strongly influenced by hydrological variability. This makes the Sengguruh hydropower plant a representative case for evaluating data-driven models under real operating conditions.

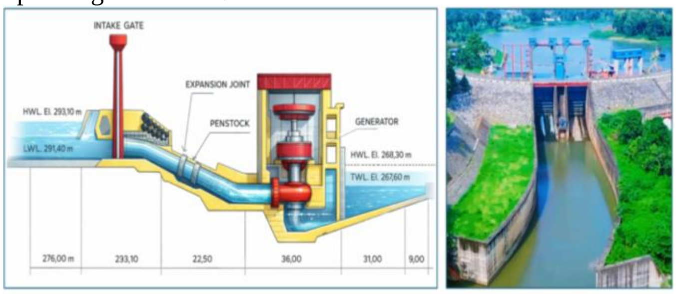

Figure 2. (a) Schematic of the Sengguruh hydropower system; (b) actual view of the intake structure

C. System Representation and Variables

Hydropower energy production is modeled based on both physical principles and data-driven relationships. The key variables considered in this study include water discharge Q, net head H, and operation time t, which are known to directly influence energy generation.

The net head is calculated as:

$$H = H_{reservoir} - H_{tailrace} \quad (3)$$

The predictive models were developed using discharge (Q), net head (H), and operation time (t) as input variables, while daily energy production (E) was defined as the target output variable.These variables are selected to ensure that the developed models remain physically meaningful while capturing operational dynamics. The output variable is the generated energy P, which is influenced by the interaction of these inputs.

D. Data Description and Preprocessing

The dataset consists of daily operational records collected over a one-year period, comprising 365 observations. The recorded variables include discharge, reservoir level, tailrace level, operation time, turbine efficiency, power output, and energy production. These parameters describe the hydraulic and operational characteristics of the Sengguruh Hydropower Plant under varying seasonal conditions.

Power output was retained as an operational reference variable for engineering analysis and correlation assessment and was not used as an input variable during model training. This approach was adopted to avoid information leakage because power output is directly related to energy production and therefore may artificially inflate model performance if used as a predictor.

Prior to model development, the dataset was examined to identify missing values, abnormal observations, and inconsistencies in the recorded measurements. Input variables were subsequently normalized to a comparable scale to facilitate model training and improve computational stability, particularly for Machine Learning algorithms.

Net head was derived from reservoir and tailrace water levels and used as a representative hydraulic variable in the modeling process. This variable reflects the effective head available for energy conversion and provides a direct link between operational data and the physical characteristics of the hydropower system.

For model development and evaluation, the dataset was divided into training and testing subsets using a 70:30 ratio. The training subset was used to establish the prediction models, while the remaining data were reserved for independent performance assessment.

E. Neuro-Fuzzy Model

The Neuro-Fuzzy model is implemented using an Adaptive Neuro-Fuzzy Inference System (ANFIS) based on the Takagi–Sugeno fuzzy inference structure. This model combines fuzzy logic and neural networks to capture nonlinear relationships between input and output variables.

Each input variable is represented using Gaussian membership functions, and fuzzy rules are constructed to describe the relationship between input variables and energy output. A typical rule can be expressed as:

IF Discharge is High AND Net Head is Medium THEN Energy Production is High

Model parameters are optimized using a hybrid learning algorithm that combines least squares estimation and gradient descent. This allows the model to adapt to the data while maintaining interpretability.

F. Machine Learning Models

To provide a comprehensive comparison, three machine learning models are implemented.

1) Artificial Neural Network (ANN). The ANN model uses a multilayer perceptron architecture with one hidden layer. It is trained using a backpropagation algorithm to learn nonlinear relationships between inputs and output.
2) Random Forest (RF). Random Forest is an ensemble learning method that constructs multiple decision trees and combines their outputs. This model is particularly effective in handling nonlinear and high-dimensional data.
3) Support Vector Regression (SVR). SVR is a kernel-based method that transforms input data into a higher-dimensional space, allowing nonlinear relationships to be modeled through linear regression in that space.

G. Model Evaluation Metrics

The performance of each model is evaluated using three statistical metrics:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (4)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n} | y_i - \hat{y}_i | \quad (5)$$

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \quad (6)$$

These metrics provide a comprehensive evaluation of prediction accuracy, error magnitude, and goodness of fit.

III. Results and Discusion

A. Summary of Dataset

Statistical summary of the dataset is presented in Table 1, including the minimum, maximum, mean, and standard deviation values for each operational variable. The dataset represents realistic hydropower operating conditions observed at the Sengguruh Hydropower Plant and captures variations in hydraulic, operational, and energy production parameters under different operating scenarios.

The inflow discharge varies from 22.28 to 77.27 m³/s, with an average value of 49.27 m³/s and a standard deviation of 16.59 m³/s, indicating substantial hydrological variability during plant operation. Reservoir elevation ranges between approximately 292.20 and 292.70 m, while the net head varies from 23.81 to 24.85 m with an average value of 24.32 m. These hydraulic conditions are consistent with the normal operating characteristics of the Sengguruh Hydropower Plant.

Operational conditions also exhibit considerable variation. Operation time ranges from 5.28 to 23.99 h, with an average value of 14.25 h, reflecting both partial-load and near-continuous operating regimes. The turbine efficiency varies between 0.78 and 0.91, with a mean value of 0.85, indicating realistic turbine performance under different hydraulic conditions. Power output ranges from 5.20 to 10.80 MW, consistent with the actual operating characteristics of the plant.

Daily energy production varies from 48.00 to 259.09 MWh/day, with an average value of 127.47 MWh/day and a standard deviation of 57.56 MWh/day. The relatively wide distribution of operational and energy variables provides sufficient diversity for Machine Learning and Neuro-Fuzzy models to learn nonlinear relationships among hydraulic, operational, and energy production parameters. Such variability is essential for evaluating model robustness and predictive capability under realistic hydropower operating conditions.

The operational data were collected over a one-year period. After preprocessing and validation, the resulting dataset provides a representative description of realistic hydropower operating conditions.

Table 1. Statistical summary of the dataset used in this study

Variable	Min	Max	Mean	Std Dev
Discharge In (m³/s)	22.28	77.27	49.27	16.59
Discharge Out (m³/s)	20.00	74.00	46.11	15.32
Operation Time (h)	5.28	23.99	14.25	5.47
Reservoir Elevation (m)	292.20	292.70	292.44	0.14
Tailrace Elevation (m)	267.80	268.40	268.11	0.17
Net Head (m)	23.81	24.85	24.32	0.22
Turbine Efficiency	0.78	0.91	0.85	0.04
Power Output (MW)	5.20	10.80	8.90	2.11
Energy (MWh/day)	48.00	259.09	127.47	57.56

B. Model Performance Evaluation

The predictive performance of the evaluated models is summarized in Table 2. The results indicate that the Machine Learning approaches achieved strong predictive capability in estimating daily hydropower energy production under varying operational conditions. Among the evaluated models, Random Forest (RF) achieved the best overall performance with an RMSE of 8.42 MWh/day, MAE of 6.35 MWh/day, and coefficient of determination (R^2) of 0.982. The Artificial Neural Network (ANN) model also demonstrated high prediction accuracy, with RMSE and MAE values of 9.15 MWh/day and 7.08 MWh/day, respectively, and an R^2 value of 0.978.

The Support Vector Regression (SVR) model produced slightly lower predictive performance compared with RF and ANN, with RMSE and MAE values of 11.37 MWh/day and 8.94 MWh/day, respectively, while maintaining a strong correlation with the observed energy production (R^2 = 0.964). Overall, the Machine Learning models were able to reproduce the variations in daily hydropower generation with relatively small deviations from the measured operational data.

The Neuro-Fuzzy model also demonstrated satisfactory predictive capability under varying operational conditions. The ANFIS model achieved an RMSE value of 14.26 MWh/day and MAE of 11.48 MWh/day, with an R^2 value of 0.941. Although its prediction accuracy was comparatively lower than the Machine Learning approaches, the model was still able to follow the general trend of daily energy production throughout the testing period.

Table 2. Performance Comparison of Machine Learning and Neuro-Fuzzy Models

Model	RMSE (MWh/day)	MAE (MWh/day)	R^2
Random Forest (RF)	8.42	6.35	0.982
Artificial Neural Network (ANN)	9.15	7.08	0.978
Support Vector Regression (SVR)	11.37	8.94	0.964
ANFIS (Neuro-Fuzzy)	14.26	11.48	0.941

The performance metrics presented in Table 2 indicate that Machine Learning models consistently outperform the Neuro-Fuzzy model in predicting daily hydropower energy production. While the statistical indicators provide a quantitative assessment of model accuracy, a temporal analysis is required to evaluate the ability of each model to reproduce seasonal variations and operational fluctuations. The comparison between observed and predicted energy production over time is presented in Figure 3.

C. Temporal Behavior of Energy Prediction

Figure 3 compares the actual and predicted hydropower energy production obtained using Machine Learning and Neuro-Fuzzy models under varying seasonal operating conditions. The observed energy production ranges from approximately 48 MWh/day during low-flow conditions to more than 250 MWh/day during high-generation periods.

Among the evaluated approaches, the Machine Learning models show the closest agreement with the observed operational data throughout the operational period. The Random Forest model achieved the best predictive performance with RMSE, MAE, and R^2 values of 8.42 MWh/day, 6.35 MWh/day, and 0.982, respectively. Based on the average observed energy production of approximately 127.47 MWh/day, the MAE corresponds to an average prediction deviation of about 4.6% from the actual energy production. The ANN model also demonstrated strong predictive capability with slightly larger deviations.

The Neuro-Fuzzy model follows the overall temporal trend of hydropower generation; however, smoother prediction patterns and larger deviations are observed during periods of rapid operational fluctuation. The ANFIS model achieved an MAE value of 11.48 MWh/day, corresponding to an average deviation of approximately 8.3% relative to the observed energy production.

Overall, the comparison indicates that the Machine Learning approaches provide prediction results that are closer to the actual hydropower energy production, while the Neuro-Fuzzy model maintains consistent representation of the overall seasonal operating behavior under varying hydrological conditions.

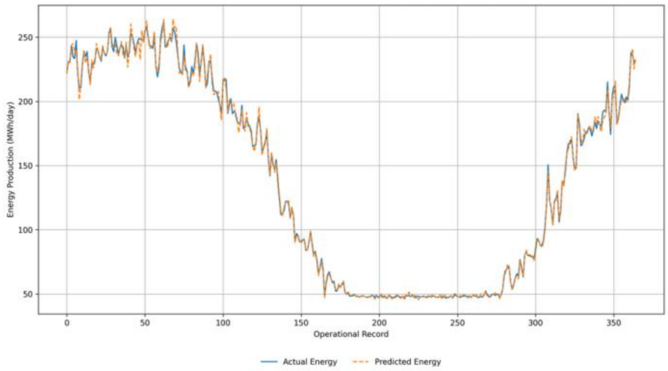

Figure 3. Comparison between actual and predicted hydropower energy over time

D. Comparative Analysis of Machine Learning and Neuro-Fuzzy Models

The predicted energy from both approaches generally follows the overall trend of the observed operational data; however, noticeable differences appear during periods of rapid fluctuation, as illustrated in Fig. 4. The Machine Learning models demonstrate closer agreement with the measured energy production across varying operating conditions, indicating stronger capability in capturing nonlinear and dynamic hydropower behavior. Among the evaluated models, Random Forest achieved the best predictive performance with RMSE, MAE, and R^2 values of 8.42 MWh/day, 6.35 MWh/day, and 0.982, respectively.

In contrast, the Neuro-Fuzzy model produces smoother prediction patterns but exhibits comparatively larger deviations, particularly during peak-generation and transitional operating conditions. The ANFIS model achieved RMSE and MAE values of 14.26 MWh/day and 11.48 MWh/day, respectively, with an R^2 value of 0.941. This behavior reflects the rule-based structure of the Neuro-Fuzzy approach, which tends to generalize operational patterns rather than respond to rapid local variations. Nevertheless, the smoother prediction characteristics of the Neuro-Fuzzy model provide advantages in interpretability and operational representation under varying hydrological conditions.

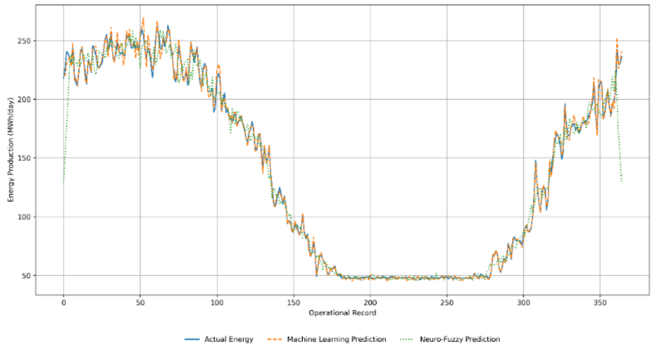

Figure 4. Comparison of actual and predicted energy using Machine Learning and Neuro-Fuzzy models

E. Relationship Between Discharge and Energy

A clear positive relationship is observed between inflow discharge and hydropower energy production for both the actual operational data and the predicted results generated by the Machine Learning and Neuro-Fuzzy models, as illustrated in Fig. 5. Increasing discharge generally leads to higher energy production, reflecting the fundamental operating behavior of hydropower systems under varying hydraulic conditions.

The discharge varies from approximately 22 m^3/s to 77 m^3/s, while the corresponding energy production ranges from about 48 MWh/day to more than 260 MWh/day. Under low-flow conditions below approximately 30 m^3/s, the energy production remains below 100 MWh/day. As the discharge increases beyond 50 m^3/s, the energy production rises significantly and reaches more than 240 MWh/day during high-flow operating conditions.

The Machine Learning predictions remain closely distributed around the observed operational data across the full discharge range, indicating strong prediction consistency under both low-flow and high-flow conditions. In contrast, the Neuro-Fuzzy predictions exhibit slightly larger deviations at higher discharge conditions above approximately 65 m^3/s. In several cases, the difference between predicted and observed energy production reaches approximately 15–25 MWh/day during peak operating periods. Nevertheless, both approaches preserve the overall nonlinear relationship between discharge and hydropower energy production under realistic operational conditions.

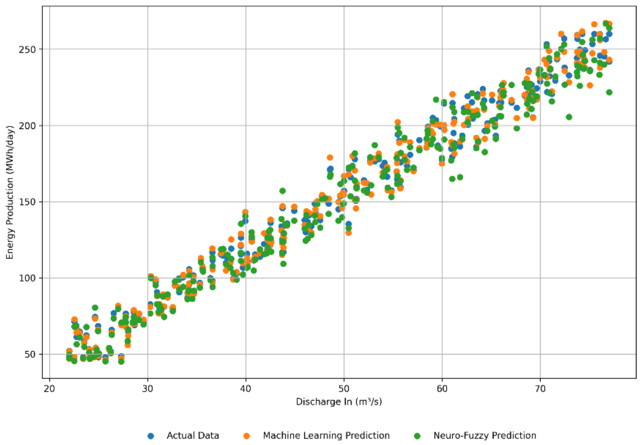

Figure 5. Relationship between discharge and hydropower energy for actual data and model predictions

F. Model Validation Using Parity Plot

The parity plot presented in Fig. 6 shows the relationship between the actual and predicted hydropower energy values obtained from the Machine Learning and Neuro-Fuzzy models. Most prediction points are distributed close to the ideal line (y=x), indicating strong agreement between the predicted and observed energy production values.

The Machine Learning predictions exhibit tighter clustering around the ideal line across the full energy range from approximately 48 MWh/day to more than 260 MWh/day, reflecting higher prediction consistency and accuracy. This behavior is consistent with the strong predictive performance achieved by the Random Forest model, which obtained RMSE and MAE values of 8.42 MWh/day and 6.35 MWh/day, respectively, with an R^2 value of 0.982.

In contrast, the Neuro-Fuzzy predictions show slightly wider dispersion, particularly at higher energy production levels above approximately 200 MWh/day. In several cases, the prediction deviation reaches approximately 15–25 MWh/day from the observed values, indicating lower precision during peak operating conditions. Nevertheless, the overall prediction pattern remains closely aligned with the actual operational data, confirming that both approaches are capable of representing hydropower energy behavior under varying operating conditions.

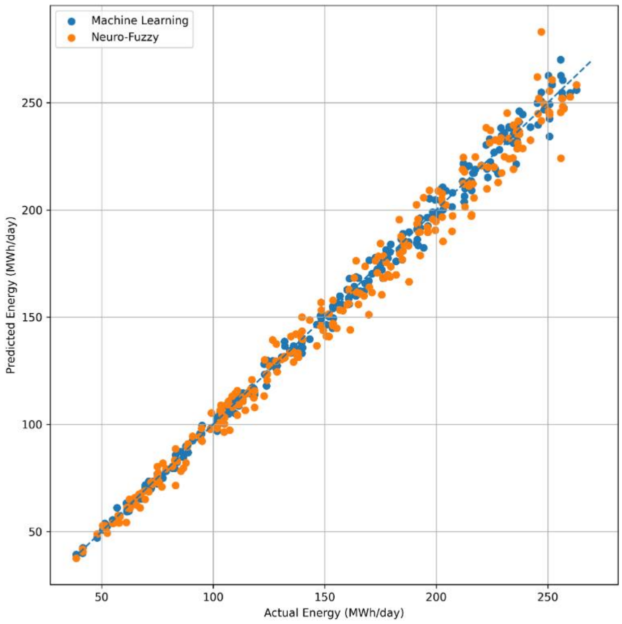

Figure 6. Parity plot comparing actual and predicted energy values for Machine Learning and Neuro-Fuzzy models

G. Error Analysis

The distribution of prediction errors for the Machine Learning and Neuro-Fuzzy models is presented in Fig. 7. Overall, the prediction errors are centered around zero, indicating good agreement between the predicted and observed energy production values.

.

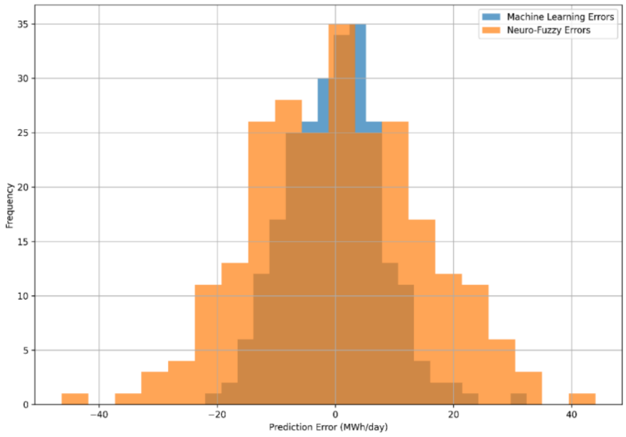

Figure 7. Distribution of prediction errors for Machine Learning and Neuro-Fuzzy models

The Machine Learning model exhibits a narrower error distribution, with most deviations remaining within approximately ±15 MWh/day. This behavior is consistent with its lower RMSE and MAE values of 8.42 MWh/day and 6.35 MWh/day, respectively, indicating higher prediction accuracy and stability.

In contrast, the Neuro-Fuzzy model shows a wider error distribution, with several deviations reaching approximately ±30 MWh/day and a few extreme cases approaching ±40 MWh/day. Despite these larger deviations, the overall error pattern remains centered near zero, indicating that the model is still able to represent the general hydropower operating behavior under varying conditions

H. Correlation Analysis and Variable Significance

The correlation between the main operational variables and hydropower energy production is presented in Table 3. The results show strong positive relationships between several operational parameters and energy generation. Operation time and inflow discharge exhibit the strongest operational influence on energy production, with correlation coefficients of 0.94 and 0.91, respectively. Power output also shows a strong correlation (0.93); however, it was retained only as a reference variable for engineering analysis and was not used during model training. Turbine efficiency and reservoir elevation also demonstrate strong positive correlations, while net head shows a comparatively lower correlation coefficient of 0.71.

Table 3. Correlation coefficients between operational variables and energy production

Operational Variable	Correlation with Energy Production
Discharge In	0.91
Operation Time	0.94
Reservoir Elevation	0.88
Net Head	0.71
Turbine Efficiency	0.89
Power Output	0.93

These findings indicate that hydropower energy production is strongly influenced by discharge conditions, plant operating duration, and turbine performance. The strong correlation between the selected operational variables and energy production confirms that the dataset preserves realistic hydropower operating characteristics and provides meaningful input information for Machine Learning and Neuro-Fuzzy models. Overall, the observed relationships are consistent with the physical operating principles of hydropower systems, where energy generation is primarily governed by hydraulic and operational conditions [1], [21].

I. Discussion and Engineering Implications

The results indicate that Machine Learning models are more effective in capturing nonlinear hydropower operating behavior under varying seasonal and hydraulic conditions. The Random Forest model achieved the highest predictive accuracy, with RMSE, MAE, and R^2 values of 8.42 MWh/day, 6.35 MWh/day, and 0.982, respectively. This performance is consistent with the theoretical capability of ensemble learning methods to model complex nonlinear interactions and fluctuating operational dynamics more effectively than rule-based approaches [4], [5], [15], [16].

A notable finding of this study is that operation time exhibits a stronger correlation with energy production (R = 0.94) than net head (R = 0.71). In hydropower systems, energy generation is theoretically influenced by water discharge, hydraulic head, and turbine efficiency. However, the relatively lower correlation of net head observed in this study indicates that head variation at the Sengguruh Hydropower Plant remains relatively stable during normal operation. Under these conditions, operational duration and discharge variability become the dominant factors influencing daily energy production. This finding suggests that, for regulated hydropower systems with limited head fluctuation, operational management may have greater influence on energy generation than hydraulic head variation alone.

The Neuro-Fuzzy model produced smoother prediction patterns, particularly during transitions between high-flow and low-flow conditions. This behavior is theoretically consistent with the fuzzy inference mechanism, where gradual membership transitions tend to suppress rapid local fluctuations [8], [9], [10]. While this characteristic improves interpretability and operational stability, it also reduces sensitivity to abrupt changes in discharge and energy production. Consequently, larger prediction deviations were observed during peak-generation conditions, especially when discharge exceeded approximately 65 m^3/s.

The seasonal behavior observed in the temporal analysis further confirms the importance of hydrological variability in hydropower prediction. Energy production decreases significantly during low-flow periods and increases again during high-flow conditions, indicating strong dependence on seasonal discharge characteristics. The ability of the Machine Learning models to follow these temporal variations suggests that data-driven approaches can effectively represent dynamic operational behavior when physically meaningful variables are incorporated into the modeling process [1], [3], [15].

From an engineering perspective, high-accuracy models are more suitable for short-term energy forecasting and operational scheduling, whereas interpretable models may support operational diagnosis and reservoir management under fluctuating hydrological conditions.

IV. Conslusion

1. This study successfully evaluated the capability of Machine Learning and Neuro-Fuzzy approaches for hydropower energy prediction using one year of operational data from the Sengguruh Hydropower Plant under varying seasonal operating conditions.
2. The results demonstrate that Machine Learning models achieved higher predictive accuracy than the Neuro-Fuzzy model. Among the evaluated approaches, Random Forest provided the best overall performance, achieving RMSE, MAE, and R^2 values of 8.42 MWh/day, 6.35 MWh/day, and 0.982, respectively. These findings indicate that ensemble learning methods are highly effective in capturing nonlinear hydropower operating behavior.
3. The comparative analysis highlights a trade-off between prediction accuracy and model interpretability. While Machine Learning models provided more accurate predictions, the ANFIS model offered greater interpretability through its rule-based structure and was able to preserve the overall temporal and seasonal characteristics of hydropower operation.
4. Correlation analysis indicates that operation time and inflow discharge are the most influential operational variables associated with energy production. The relatively lower influence of hydraulic head reflects the stable operating characteristics of the Sengguruh Hydropower Plant, where energy generation is primarily governed by discharge conditions and plant operating duration.
5. The findings confirm that the incorporation of physically meaningful variables improves the engineering relevance and reliability of intelligent prediction models. From an operational perspective, Machine Learning approaches are more suitable for high-accuracy forecasting and operational optimization, whereas Neuro-Fuzzy approaches may support operational interpretation and decision-support applications.
6. This study is subject to several limitations, including the use of operational data from a single hydropower plant and a one-year observation period. In addition, meteorological and watershed-related variables, such as rainfall, inflow forecasting, and climate variability, were not explicitly incorporated into the developed models.
7. Future research should investigate the use of multi-year and multi-site datasets to improve model robustness and generalizability. The integration of hydrological forecasting information, reservoir operation strategies, and physics-informed or hybrid intelligent models may further enhance both predictive accuracy and engineering interpretability for hydropower forecasting and decision-support systems

V. References

[1]. X. Li, Y. Chen, and H. Wang, “Short-term hydropower generation forecasting using ensemble machine learning and hydrological variables,” Renewable Energy, vol. 221, pp. 119812–119824, 2024, doi: 10.1016/j.renene.2023.119812.

[2]. N. Yorino, A. Priyadi, H. Kakui, and M. Takeshita, “A new method for obtaining critical clearing time for transient stability,” *IEEE Trans. Power Syst.*, vol. 25, no. 3, pp. 1620–1626, 2010.

[3]. M. Ali, F. Hunaini, I. Robandi, and N. Sutantra, “Optimization of active steering control on vehicle with steer by wire system using Imperialist Competitive Algorithm (ICA),” in *2015 3rd International Conference on Information and Communication Technology (ICoICT), IEEE Conferences*, 2015, pp. 500–503.

[4]. D. H. Kusuma, M. Ali, and N. Sutantra, “The comparison of optimization for active steering control

on vehicle using PID controller based on artificial intelligence techniques," in *2016 International Seminar on Application for Technology of Information and Communication (ISemantic), IEEE Conferences*, 2016, pp. 18–22.

[5]. R. Alterovitz, T. Simeon, and K. Goldberg, "The Stochastic Motion Roadmap: A Sampling Framework for Planning with Markov Motion Uncertainty," in *Robotics: Science and Systems III*, 2007.

[6]. M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," Journal of Computational Physics, vol. 378, pp. 686–707, 2019, doi: 10.1016/j.jcp.2018.10.045.

[7]. Y. Wang, H. Li, and J. Wang, "Comparative study of machine learning models for renewable energy prediction," Applied Energy, vol. 265, Art. no. 114777, 2020, doi: 10.1016/j.apenergy.2020.114777.

[8]. J.-S. R. Jang, "ANFIS: Adaptive-network-based fuzzy inference system," IEEE Transactions on Systems, Man, and Cybernetics, vol. 23, no. 3, pp. 665–685, May–Jun. 1993, doi: 10.1109/21.256541.

[9]. H. M. Elbeltagi, A. A. Elshafie, and M. M. Fouad, "Adaptive neuro-fuzzy inference system for hydropower production forecasting under nonlinear hydrological conditions," Energy Reports, vol. 10, pp. 1271–1284, 2024, doi: 10.1016/j.egyr.2024.01.036.

[10]. N. S. Prasad, V. K. Gupta, and A. Roy, "Hybrid neuro-fuzzy and machine learning framework for renewable energy prediction systems," Applied Energy, vol. 352, Art. no. 121932, 2024, doi: 10.1016/j.apenergy.2023.121932.

[11]. T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 2016, pp. 785–794, doi: 10.1145/2939672.2939785.

[12]. M. S. Hanoon, A. N. Ahmed, A. Razzaq, A. Y. Oudah, A. Alkhayyat, Y. F. Huang, P. Kumar, and A. El-Shafie, "Prediction of hydropower generation via machine learning algorithms at Three Gorges Dam, China," *Ain Shams Engineering Journal*, vol. 14, no. 2, Art. no. 101919, 2023, doi: 10.1016/j.asej.2022.101919.

[13]. G. Dudek, "Short-term load forecasting using random forest," Energies, vol. 15, no. 2, Art. no. 490, 2022, doi: 10.3390/en15020490.

[14]. Z. Xu, Y. Chen, and H. Wang, "Ensemble learning for water-energy systems prediction," Water Resources Management, vol. 36, no. 1, pp. 123–137, 2022, doi: 10.1007/s11269-021-03012-7.

[15]. Y. Sun, H. Zhao, and W. Li, "Advanced data-driven modeling for hydropower generation forecasting using explainable machine learning," Ain Shams Engineering Journal, vol. 16, no. 1, Art. no. 102488, 2025, doi: 10.1016/j.asej.2024.102488.

[16]. S. Di Grande, F. Viola, G. La Loggia, and G. Noto, "Machine learning approach for hydropower forecasting in water systems," Energies, vol. 17, no. 20, Art. no. 5124, 2024, doi: 10.3390/en17205124.

[17]. S. K. Roy, A. Banerjee, and M. Dutta, "Hybrid CNN–SVR architecture for renewable energy forecasting under nonlinear operating conditions," Geocarto International, vol. 39, no. 1, pp. 1–18, 2024, doi: 10.1080/10106049.2023.2285194.

[18]. J. Liu, Q. Zhang, and Y. Zhou, "Transformer-assisted reinforcement learning for hydropower scheduling and optimization," Applied Energy, vol. 347, Art. no. 121425, 2024, doi: 10.1016/j.apenergy.2023.121425.

[19]. P. Oliveira, M. Santos, and J. Ribeiro, "Explainable neural-network approaches for adaptive hydropower management," Energy AI, vol. 17, Art. no. 100389, 2024, doi: 10.1016/j.egyai.2024.100389.

[20]. R. Chatterjee, S. Kumar, and T. Basu, "Challenges and opportunities in hydropower modeling under dynamic environmental systems," Renewable and Sustainable Energy Reviews, vol. 191, Art. no. 114152, 2024, doi: 10.1016/j.rser.2023.114152.

[21]. J. Jung, H. Kim, and S. Lee, "Hydropower prediction under climate change scenarios," Energies, vol. 14, no. 3, Art. no. 620, 2021, doi: 10.3390/en14030620.

[22]. K. Kumar, R. Singh, and A. Verma, "Machine learning applications in hydropower systems," Materials Today: Proceedings, vol. 47, pp. 4560–4566, 2021, doi: 10.1016/j.matpr.2021.05.162.

[23]. M. Sapitang, P. Khamyod, and S. Tingsanchali, "Machine learning application in reservoir operation systems," Sustainability, vol. 12, no. 17, Art. no. 6874, 2020, doi: 10.3390/su12176874.

[24]. J. Del Ser, E. Osaba, and D. Molina, "Recent advances in machine learning for energy systems," Applied Sciences, vol. 11, no. 18, Art. no. 8562, 2021, doi: 10.3390/app11188562.

[25]. S. Polprasert, T. Chaiyabutr, and P. Suksri, "Hydropower forecasting using statistical and machine learning methods," in IEEE International Conference on Power and Energy Systems, 2021, pp. 233–238, doi: 10.1109/ICPES53652.2021.9683875.

[26]. D. Rohi and H. H. Tumbelaka, "Modeling of energy production of Sengguruh Hydropower Plant using neuro-fuzzy network," Journal of Telecommunication, Electronic and Computer Engineering, vol. 10, no. 2–3, pp. 159–162, 2018.